Nucleic Acids Research, 1995, Vol. 23, No. 15 2815–2822

A novel human c-*sis* mRNA species is transcribed from a promoter in c-*sis* intron 1 and contains the code for an alternative PDGF B-like protein

Ron P. H. Dirks*, Carla Onnekink, Hans J. Jansen, Aard de Jong and Henri P. J. Bloemers

Department of Biochemistry, University of Nijmegen, PO Box 9101, 6500 HB Nijmegen, The Netherlands

Received June 1, 1995; Accepted June 26, 1995

EMBL accession nos X83705 and X83706

ABSTRACT

The human platelet-derived growth factor (PDGF) B chain precursor is usually translated from a 3.5 kb c-*sis*/PDGF B gene transcript. The first exon of the c-*sis* gene contains the code for the signal peptide of the PDGF B chain precursor, preceded by a 1 kb long untranslated sequence with potent translation inhibitory activity. In this paper we show that a novel 2.6 kb c-*sis* mRNA present in the human choriocarcinoma cell line JEG-3 initiates at an alternative exon 1, which we refer to as exon 1a. The 90 bp long exon 1a is located in the center of the first intron of the gene. It coincides with a very pronounced DNase-I-hypersensitive site and is preceded by a functional promoter. Of the three ATG codons present in exon 1a, the third one perfectly matches the criteria of a consensus start codon. It initiates an open reading frame that is continuous with the code for the PDGF B chain precursor but lacks the code for a signal peptide. We conclude that this novel 2.6 kb c-*sis* mRNA species lacks the strong translation inhibitory potential of the regular exon 1 and contains the code for a PDGF B-like protein that may be targeted to the cell nucleus.

INTRODUCTION

The human c-*sis* gene encodes the precursor of the platelet-derived growth factor (PDGF) B chain (1–5), a potent growth factor for cells of mesenchymal origin. The gene is located on human chromosome 22 (6,7). It consists of seven exons, which span ~22 kb (8,9). Together the exons form an mRNA of 3.5 kb, of which 2.8 kb consist of non-translated sequences. Exon 1 contains a 1 kb long 5′-untranslated sequence with a strong translation inhibitory potential (10–12) and, in addition, codes for the signal peptide. Exons 2–6 code for the major part of the PDGF B chain precursor, whereas exon 7 consists entirely of a 1.6 kb long 3′-untranslated sequence (8,9). In the endoplasmic reticulum (ER), the 28 kDa PDGF B chain precursor undergoes N-linked glycosylation and dimerizes with other B and/or A chain precursors (13–17). The B chain precursors are proteolytically processed in the Golgi system at their N- and C-termini, which results in the loss of the N-linked carbohydrate. Finally, PDGF B chains are secreted as part of 30 kDa AB or BB dimers, whereas 24 kDa BB dimers remain cell associated and are degraded in lysosomes (16,17).

The human c-*sis* gene is transcribed in specific cell types, such as vascular endothelial cells (18), macrophages (19), activated monocytes (20), placental cytotrophoblasts (21) and bone marrow megakaryocytes (22). Furthermore, the gene is transcribed in several types of transformed cells (e.g. 23,24). The full length 3.5 kb c-*sis* transcript is found in most human cell types expressing the gene, but smaller transcripts have been described that hybridize under stringent conditions with c-*sis* or v-*sis* probes (18,23, 25–29). Among these, an ~2.6 kb long transcript has been described most often. Fen and Daniel (29) provided evidence that a 2.8 kb c-*sis* transcript expressed in human umbilical vein endothelial cells (HUVEC) lacks a major part of c-*sis* exon 1 and initiates at 15 nucleotides upstream of the PDGF B translation initiation codon. Since the code for the PDGF B chain precursor is presumably not affected in the 2.8 kb transcript, the main difference compared with the 3.5 kb mRNA would be the absence of the translation inhibitory capacity of the 5′-untranslated region (10–12). Therefore, it was predicted that the 2.8 kb transcript would be much more efficiently translated than the 3.5 kb mRNA (29).

A 2.6 kb c-*sis* mRNA species expressed in human choriocarcinoma cell line JEG-3 was also shown to lack exon 1-derived sequences and was suggested to be antisense to the major 3.5 kb c-*sis* mRNA (28). Indeed, in JEG-3 cells the regular c-*sis* promoter does not coincide with a DNase-I-hypersensitivity (DH) site (30), which strongly suggests that the promoter is not active in these cells. However, the presence of a DH site at the 3′-end of the first exon in JEG-3 cells (30) suggests that the 2.6 kb c-*sis* transcript could be derived from an alternative promoter, corresponding with the start site of the 2.8 kb c-*sis* mRNA in vascular endothelial cells (29). When we performed primer extension and nuclease S1 experiments, similar to those described by Fen and Daniel (29), on RNA isolated from JEG-3 cells, we found that the 2.6 kb c-*sis* mRNA has the same orientation as the 3.5 kb mRNA and that its 5′-end extends 90 nucleotides (nt) upstream of exon 2-derived sequences. Surprisingly, the 5′-end of the 2.6 kb transcript turned out to be different from the 3′-end of the regular c-*sis* exon 1. These results prompted us to map the 2.6 kb c-*sis* mRNA from JEG-3 cells in more detail, which led to the discovery that the transcript initiates at an alternative exon

* To whom correspondence should be addressed at present address: Department of Molecular Biology, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands

located within c-*sis* intron 1. Our results indicate that in JEG-3 cells the c-*sis* gene uses a hitherto unknown expression pathway that may not only affect the translation efficiency of the c-*sis* mRNA, but also the structure and cellular location of its protein product.

MATERIALS AND METHODS

Cell culture

Cell lines K562, JEG-3, JAR, HeLa and T24 were all from the American Type Culture Collection (Rockville, MD). Cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum and 125 U/ml streptomycin and 125 U/ml penicillin. Megakaryocytic differentiation of K562 cells was induced by addition of 12-*O*-tetradecanoylphorbol 13-acetate (TPA; Sigma) to the culture medium (final concentration, 2 ng/ml).

Isolation and Northern blot analysis of RNA

Total cellular RNA was isolated according to the LiCl/urea method (31). RNA was glyoxylated, electrophoresed in 1% agarose gels containing 10 mM Na_2HPO_4, pH 7.0 and transferred to Hybond-N filters (Amersham, UK) as described (32). RNA was hybridized overnight in 0.5 M Na_2HPO_4, pH 7.2, 1% bovine serum albumin, 7% SDS, 1 mM EDTA at 65°C with ^{32}P-labeled DNA probes. Filters were washed at 65°C in 0.1 M Na_2HPO_4, 0.5% SDS, 1 mM EDTA and exposed to Kodak XAR-5 X-Ray film with two Cronex intensifying screens (Du Pont).

Plasmids and probes

A 2 kb *Bam*HI fragment that contains sequences from c-*sis* exons 1–7 was isolated from human c-*sis* cDNA clone pSM-1, which was kindly provided by Dr S. T. Josephs. A 2 kb *Eco*RI–*Hin*dIII fragment containing all c-*sis* exon 1-derived sequences was isolated from pAO121. A 1.45 kb *Bam*HI fragment that contains the 3′ half of exon 7 was isolated from human c-*sis* cosmid clone ALLW-1283-Cl21. The 0.6 kb *Pst*I–*Bgl*II fragment of pSM-1 was cloned into pGEM-3Zf(+) and was used as the probe for nuclease S1 analysis. The 0.35 kb *Pst*I–*Kpn*I fragment of pAO78 was used as a probe for DH site mapping. pAO78 contains the 8 kb *Bam*HI fragment of ALLW-1283-Cl21 cloned into pBR322. pAO152 contains the 2 kb *Hin*dIII–*Bam*HI fragment of ALLW-1283-Cl21 cloned into pAT153. Genomic c-*sis* clones pAO70, pAO121 and ALLW-1283-Cl21 (33) and pAO154 (34) have been described earlier. Human PDGF A cDNA clone 13.1 was kindly provided by Dr C. Betsholtz. All DNA probes used in Southern blot and Northern blot analyses were labeled with ^{32}P by the random primer labeling method (35) to specific activities of 10^8–10^9 c.p.m./μg DNA.

The reporter gene construct psis–112/+43CAT has been described earlier (30). pSuperCAT is a promoterless chloramphenicol acetyltransferase (CAT) gene vector. pAltsis(s)CAT and pAltsis(a)CAT were made by cloning the 0.22 kb *Dde*I fragment of pAO78 in the sense (s) or antisense (a) orientation into the *Sma*I site of pSuperCAT. pSV2CAT was described earlier (36). pCH110 (Pharmacia LKB) contains the β-galactosidase gene driven by the SV40 promoter.

Primer extension analysis

cDNA was synthesized from 30 μg JEG-3 and JAR RNA using the ^{32}P-end-labeled exon 2-derived antisense oligonucleotide 5′-(GTG CAG CAG GCG TTG GAG ATC ATC AAA GGA GCG GA)-3′ as a primer, essentially as was described by Fen and Daniel (29). Primer extension products were electrophoresed in a 6% sequencing gel and exposed to Kodak XAR5 X-ray film for 2 weeks with two intensifying screens. The same exon 2-derived oligonucleotide was used as a primer to generate a nucleotide sequence ladder (37) of c-*sis* cDNA clone pSM-1, which served as a marker for the primer extension analysis.

Nuclease S1 analysis

Nuclease S1 analysis was done as described earlier (38). The nuclease S1 probe (see Fig. 3A) was labeled with ^{32}P according to Sanger *et al.* (37) and purified by electrophoresis in a 6% sequencing gel. RNA from TPA treated K562 cells (50 μg) and JEG-3 cells (250 μg) was hybridized overnight at 40°C with ^{32}P-labeled probe. Treatment with nuclease S1 (final concentration 2 U/μl) was for 60 min at 37°C. Protected fragments were electrophoresed in a 6% sequencing gel and exposed to Kodak XAR-5 film. The dideoxy nucleotide sequence ladder of the S1 probe (37) was used as a molecular weight marker.

cDNA cloning

Polyadenylated RNA from JEG-3 cells was purified by oligo(dT)-cellulose affinity chromatography. cDNA synthesis and amplification were according to the protocol of the 5′-AmpliFINDER RACE Kit (Clontech), which is based on the single-strand-ligation-to-single-stranded-cDNA (SLIC) method according to Dumas *et al.* (39). cDNA was synthesized from 2 μg polyadenylated RNA using c-*sis* exon 6-specific antisense oligonucleotide 5′-(GTC ACC CGA GTT TGG GGC GT)-3′ as a primer. Upon ligation of the AmpliFINDER Anchor to its 3′ end, the cDNA was amplified using an oligonucleotide complementary to the AmpliFINDER Anchor Primer and a nested c-*sis* exon 5-specific antisense oligonucleotide 5′-(CTC GCT GCT CCT GGG AAC CC)-3′. The cDNA was amplified in 35 cycles: denaturation was for 45 s at 94°C, annealing for 45 s at 60°C and synthesis for 2 min at 72°C. The resulting 600 bp cDNA fragment was cloned into the *Sma*I site of pGEM-3Zf(+), yielding pCORD1.

Nucleotide sequence analysis

DNA fragments were subcloned into pGEM-3Zf(+) and nucleotide sequences were determined according to the dideoxy chain termination method of Sanger *et al.* (37). Sequence data were recorded, edited and compared using IntelliGenetics Suite software. The cDNA sequence of pCORD1 (EMBL accession no. X83705) and the genomic DNA sequence of the alternative exon 1a (accession no. X83706) were submitted to the EMBL data library.

Mapping of DH sites and Southern blot analysis

Nuclear isolations, DNase I digestions and chromosomal DNA purifications were all performed as previously described (30). Purified DNA (25 μg) was digested to completion with *Kpn*I (Gibco Laboratories, Grand Island, USA), extracted once with

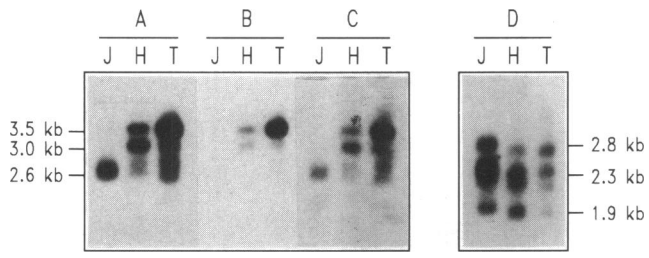


Figure 1. Northern blot analysis of c-*sis* mRNA species. Total RNA isolated from JEG-3 (J), HeLa (H) and T24 (T) cells was hybridized with cDNA clone pSM-1 containing sequences derived from all seven c-*sis* exons (**A**), with genomic clone pAO121 containing the entire c-*sis* exon 1 (**B**), with a 1.45 kb genomic fragment containing the 3′ end of c-*sis* exon 7 (**C**) and with PDGF A cDNA clone 13.1 (**D**). JEG-3 and HeLa lanes contain 10 µg of RNA; T24 lanes contain 5 µg of RNA.

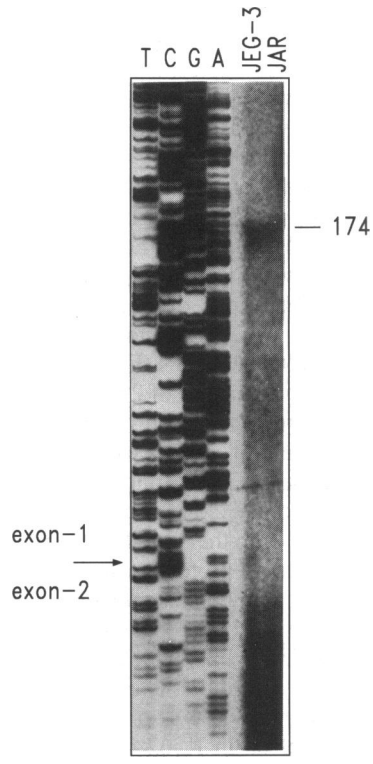


Figure 2. Primer extension analysis of the 2.6 kb c-*sis* mRNA. Total RNA derived from choriocarcinoma cell lines JEG-3 and JAR was hybridized with a c-*sis* exon 2-specific oligonucleotide and the primer was extended using reverse transcriptase. The size of the single extension product is indicated. The sequence ladder was generated by dideoxy nucleotide sequence analysis of c-*sis* clone pSM-1 using the same exon 2-specific primer. The boundary between exon 2- and exon 1-derived sequences is indicated by the arrow on the left.

phenol and chloroform, precipitated with ethanol and dissolved in 20 µl of H_2O. The *Hin*dIII fragments of phage λ (400 ng/lane) and the *Hae*III fragments of phage ΦX174 (200 ng/lane) were added to each sample as internal molecular size markers. DNA samples were electrophoresed in 1% agarose gels and transferred to Hybond-N filters according to the Amersham protocol. DNA was hybridized overnight in 0.75 M NaCl, 75 mM sodium citrate, 0.1% SDS, 0.1% Ficoll, 0.1% bovine serum albumin, 0.1% polyvinylpyrrolidone, 100 µg/ml herring sperm DNA at 68°C with ^{32}P-labeled DNA probes. Filters were washed at 68°C in 15 mM NaCl, 1.5 mM sodium citrate, 0.1% SDS, then exposed to X-ray film with two intensifying screens. For genomic localization of the alternative exon 1a, subcloned c-*sis* fragments were isolated according to standard protocols (32) and electrophoresed in a 1% agarose gel. Southern blot analysis was performed as described above using the ^{32}P-labeled 60 bp *Eco*RI–*Ava*I fragment of c-*sis* cDNA clone pCORD1 as a probe.

Electropermeations and CAT assays

Supercoiled plasmid DNA was purified by CsCl gradient centrifugation (32) and quantitated by means of both ethidium bromide staining and spectrophotometric measurement. Cells ($3–5 \times 10^7$) were electropermeated with 10 µg pCH110 and 10 pmol CAT construct as previously described (30). Cells were harvested 48 h after electropermeation. Cell lysates were prepared (32) and the protein content of each lysate was determined in a protein assay (BioRad). Fixed amounts of protein were tested for β-galactosidase activity (40) and, subsequently, amounts of protein corresponding to equal β-galactosidase activity were assayed for CAT activity (36). Radioactive spots were quantitated by measuring in a liquid scintillation counter, from which the percentage acetylated chloramphenicol could be determined.

RESULTS

The 5′-end of a 2.6 kb c-*sis* mRNA species is derived from an alternative exon

A representative Northern blot analysis, using a probe that contains sequences derived from all seven c-*sis* exons, demonstrates the cell-type specific expression of three c-*sis* mRNA species (Fig. 1A). A single 2.6 kb c-*sis* mRNA species is expressed in choriocarcinoma-derived JEG-3 cells, whereas bladder carcinoma-derived T24 cells express the 3.5 kb, a 3.0 kb and a 2.6 kb transcript. Cervix carcinoma-derived HeLa cells express the 3.5 kb and a 3.0 kb species. A parallel Northern blot analysis using a genomic probe that contains only c-*sis* exon 1 (Fig. 1B) shows that the 2.6 kb transcript does not hybridize with exon 1-derived sequences. The 3.0 kb transcript hybridizes with the exon 1 probe, but not as well as with the cDNA clone pSM-1 (Fig. 1A and B), which suggests that it lacks a significant part of exon 1. The hybridization signals obtained with a probe that contains only sequences derived from the 3′-end of the 1.6 kb long exon 7 (Fig. 1C) are comparable with those obtained with pSM-1 (Fig. 1A), suggesting that the different sizes of the three transcripts do not result from alternative polyadenylation. Northern blot analysis with additional exon-specific probes indicated that the 2.6 kb transcript also contains sequences derived from c-*sis* exons 5 and 6 and from the 5′-end of exon 7 (data not shown). We conclude that the 3.0 kb c-*sis* mRNA expressed in HeLa and T24 cells initiates within c-*sis* exon 1 and may in fact be identical to a 3.0 kb mRNA species expressed in vascular endothelial cells (29). The 2.6 kb mRNA species seems to lack exon 1-derived sequences, which is in agreement with the results described by Franklin *et al.* (28). However, as was demonstrated for the 2.8 kb c-*sis* mRNA species expressed in vascular endothelial cells (29), the 2.6 kb transcript might contain

insufficient exon 1-derived sequences to allow visualization by Northern blot analysis with an exon 1-specific probe. Since JEG-3, HeLa and T24 cells express significant levels of PDGF A mRNA (Fig. 1D), it was even possible that the hybridization signal at 2.6 kb was caused by a PDGF A transcript cross-hybridizing with the c-*sis* probe.

To examine whether the 2.6 kb mRNA expressed in JEG-3 cells starts at the 3′-end of c-*sis* exon 1, as has been demonstrated for the 2.8 kb c-*sis* mRNA expressed in endothelial cells (29), the 5′-end of the transcript was mapped by primer extension analysis, using the exon 2-derived primer described by Fen and Daniel (29). A single 174 nt primer extension product was found for JEG-3 cells as well as for JAR cells (Fig. 2), which also express only a single c-*sis* mRNA species of 2.6 kb (30,41). The size of the primer extension product indicates that the 2.6 kb transcript extends 90 nt upstream of c-*sis* exon 2-derived sequences. Together with c-*sis* exons 2–7 and a poly A tail of 200 nt this fully accounts for the length of the 2.6 kb transcript. To check whether these 90 nt are derived from the 3′-end of the regular c-*sis* exon 1, we performed nuclease S1 protection analysis on RNA derived from JEG-3 cells using a probe that extended from the 3′-end of exon 1 to the 5′-end of exon 4 (as depicted in Fig. 3A). RNA from phorbol ester-treated K562 cells, which express only the 3.5 kb transcript, was used as a control. As expected, all c-*sis*-derived sequences present in the nuclease S1 probe were protected by the 3.5 kb transcript expressed in K562 cells, which resulted in a 435 nt long fragment (Fig. 3B). Surprisingly, the nuclease S1 probe protected only a 255 nt long sequence in RNA from JEG-3 cells (Fig. 3B), instead of a 345 nt long region that was expected if the 2.6 kb mRNA would start at the 3′-end of exon 1. In fact, in JEG-3 RNA, the protected region initiates exactly at the boundary between exons 1 and 2. Thus, the combined results of the Northern blot, primer extension and nuclease S1 experiments indicate that the choriocarcinoma cell line JEG-3 expresses a 2.6 kb c-*sis* mRNA species that extends 90 nt upstream of c-*sis* exon 2-derived sequences and initiates at an alternative exon.

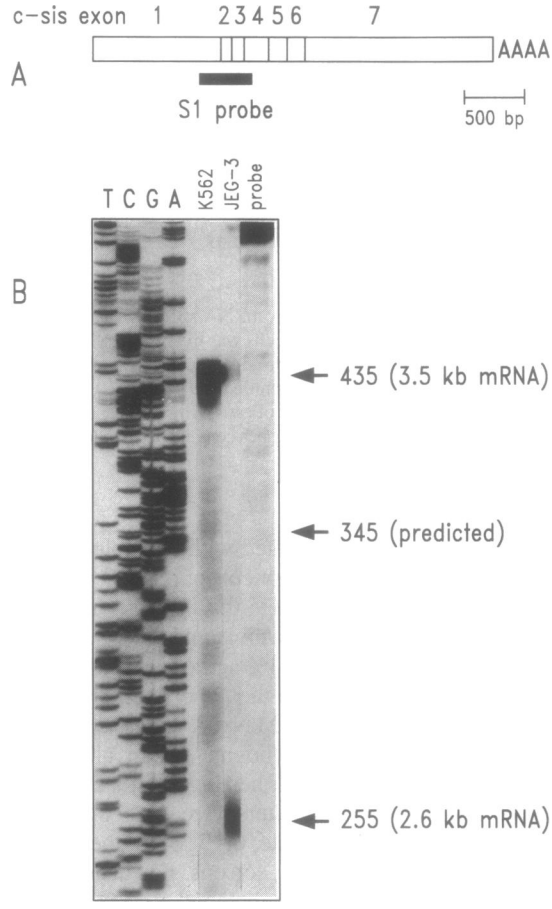


Figure 3. Nuclease S1 analysis of the 2.6 kb c-*sis* mRNA. (**A**) Schematic representation of the probe that was used for the nuclease S1 analysis. c-*sis* exons 1–7 are boxed. The probe is indicated by the black bar. (**B**) Total RNA derived from TPA-treated K562 cells and JEG-3 cells was hybridized with the ^{32}P-labeled nuclease S1 probe (Fig. 3A). The RNA–DNA hybrids were treated with nuclease S1 and the protected fragments were analysed by electrophoresis in a sequencing gel. The sizes of the protected fragments were determined by comparison with a parallel sequence ladder generated by dideoxy nucleotide sequence analysis of the nuclease S1 probe. Arrows indicate the positions of the protected fragments and the position of the predicted 345 nt fragment.

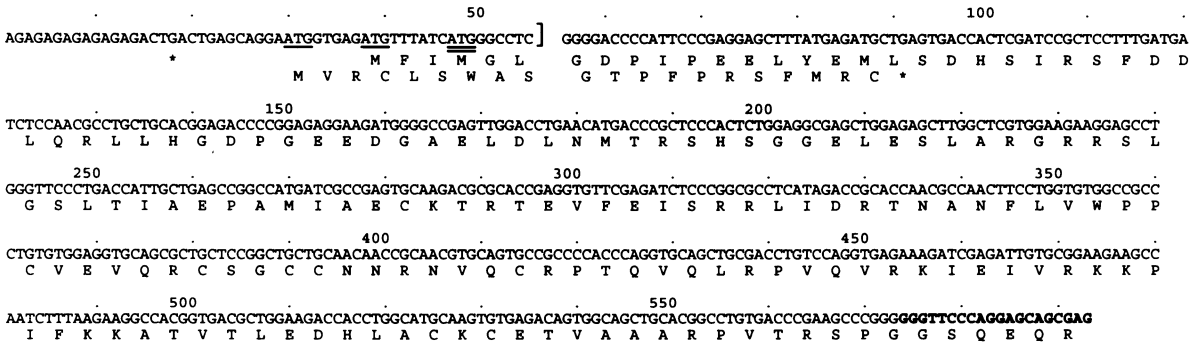
The 2.6 kb c-*sis* mRNA species contains the code for a novel PDGF-B-like protein

In order to identify the alternative c-*sis* exon, we decided to isolate a cDNA corresponding to the 5′-end of the 2.6 kb transcript by a cDNA amplification method (single-strand-ligation-of-cDNA or SLIC-PCR; 39). First strand cDNA was synthesized from JEG-3 RNA using a c-*sis* exon 6-derived primer. A single stranded adapter was ligated to the 3′-end and the cDNA was amplified in a polymerase chain reaction using a nested c-*sis* exon 5-derived primer and an oligonucleotide complementary to the adapter. The resulting cDNA was cloned into a pGEM vector yielding clone pCORD1 and its nucleotide sequence was determined (Fig. 4A). The cDNA clone contains the previously described c-*sis* exons 2–5 (3–5) and, in addition, a novel 55 bp sequence located immediately upstream of exon 2. The novel sequence contains three ATG codons, two of which are in frame with the code for the PDGF B chain precursor that continues in exon 2. The first ATG is followed by a stop codon in exon 2 and initiates an open reading frame of only 20 codons. The third ATG codon perfectly matches the criteria of a consensus translation initiation site according to Kozak (42; see Fig. 4B). Translation from this codon would result in the synthesis of a PDGF B-like polypeptide that contains the tripeptide met-gly-leu at its amino-terminus, followed by amino acids that are encoded by exon 2. As a result, the signal peptide is replaced by an N-terminus harbouring three negatively charged residues within the first ten amino acids.

The alternative c-*sis* exon 1a is preceded by a functional promoter and coincides with a pronounced DH region in c-*sis* intron 1

Earlier, we and others localized very pronounced DH sites within c-*sis* intron 1 in JEG-3 cells (28,30). When we used the novel 55 bp cDNA fragment as a probe for Southern blot analysis of subcloned genomic c-*sis* fragments (Fig. 5A), c-*sis* intron 1 (subclone pAO70) was found to hybridize under highly stringent conditions (Fig. 5B). Ultimately, the hybridizing region could be localized on a 1.4 kb *Pvu*II fragment (Fig. 5C), of which the nucleotide sequence was partly determined (Fig. 6). Indeed, the *Pvu*II fragment contains the same 55 bp sequence as is present in the 5′-end of c-*sis* cDNA clone pCORD1. The 55 bp sequence is

A



B

	-3		+4
first Met codon	GGA	ATG	G
second Met codon	GAG	ATG	T
third Met codon	ATC	ATG	G
consensus startcodon	ACC	ATG	G

Figure 4. Nucleotide sequence and deduced amino acid sequence of c-*sis* cDNA clone pCORD1. (**A**) The 3′ boundary of the novel 55 bp long sequence is indicated by parentheses. ATG codons in the novel sequence are underlined. The consensus translation initiation site is double underlined. Translation termination codons are indicated by asterisks. The sequence complementary to the exon 5-derived oligonucleotide used in the SLIC-PCR procedure is shown in bold. (**B**) Alignment of the ATG codons in the novel 55 bp sequence with the consensus translation initiation site according to Kozak (42).

followed by a consensus 5′-splice site. The 5′-end of the alternative exon, which we refer to as exon 1a, was localized at 90 bp upstream of its 3′-end, according to the results of the primer extension experiment (Fig. 2). The presumptive transcription start site of the 2.6 kb transcript is not preceded by a TATA box and maps at a remarkable sequence, where a polypurine stretch $(dGdA)_7$ switches to an alternating purine/pyrimidine stretch $(dGdPy)_{11}$.

To test whether exon 1a is preceded by a functional promoter, a genomic *Dde*I fragment containing the region between positions −191 to +28 relative to the presumptive transcription start site (see also Fig. 6) was cloned in front of the chloramphenicol acetyltransferase (CAT) gene and assayed for promoter activity in JEG-3 cells (Fig. 7). In the sense orientation (pAltsis(s)CAT), the 0.22 kb *Dde*I fragment has activity that is ~15-fold lower than the activity of the regular c-*sis* promoter (psis −112/+43CAT), whereas in the antisense orientation (pAltsis(a) CAT) it displays only background activity.

To examine whether the alternative exon 1a coincides with a DH site previously found in JEG-3 cells (28,30), we decided to map the boundaries of this DH site more precisely. By Southern blot analysis of a 3.3 kb *Kpn*I fragment we could localize the DH site between 1.9 and 1.3 kb upstream of the 3′-*Kpn*I site (Fig. 8), which corresponds exactly with the positions of the alternative exon 1a and its immediate upstream region (see also Fig. 5A). We could also detect two small protected regions within the DH site, which may result from the binding of specific *trans*-acting factors. Our results indicate that the alternative exon 1a is preceded by a functional promoter and coincides exactly with a DH site in intron 1, at ~4 kb downstream of the transcription start site of the regular 3.5 kb c-*sis* mRNA.

DISCUSSION

The human c-*sis*/PDGF B gene uses several different ways to regulate its expression. Transcription of the gene is regulated in a cell type- and developmental stage-specific manner and can be influenced by several extracellular agents (reviewed in 43,44). Translation can be inhibited up to 40-fold by the 1 kb long GC-rich leader of the 3.5 kb c-*sis* mRNA (10–12). Expression of a 2.8 kb c-*sis* mRNA that initiates at the 3′ end of exon 1 may provide a means to escape from the translation inhibitory effect of the leader sequence (29). The 3.5 kb c-*sis* mRNA has a variable stability depending on the cell type in which it is expressed (e.g. 45,46). All these regulatory mechanisms have in common that they modulate the expression level of the PDGF B gene, but do not affect the code for the PDGF B chain precursor. We now demonstrate a completely new expression pathway for the human c-*sis* gene: initiation of transcription at an alternative promoter located in intron 1, at 4 kb downstream of the transcription start site of the 3.5 kb transcript, yields a 2.6 kb c-*sis* mRNA that lacks the GC-rich leader. Sequence analysis of a cDNA that corre-

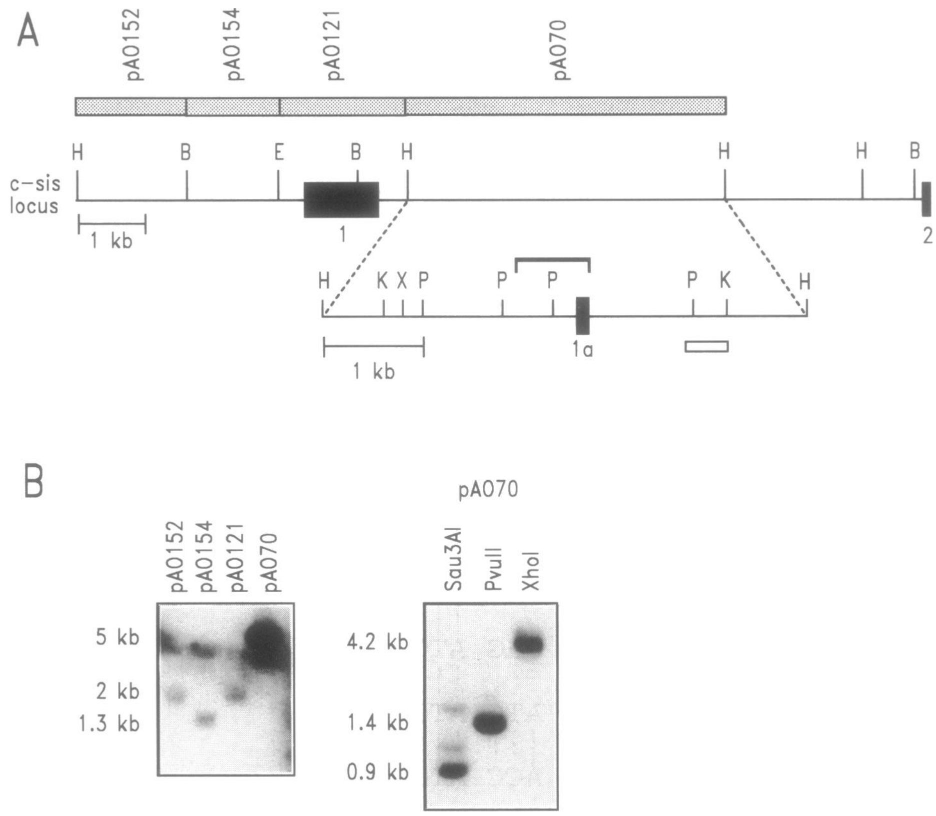


Figure 5. Localization of the alternative c-*sis* exon 1a. (**A**) Restriction enzyme map of c-*sis* intron 1 and flanking regions. Genomic subclones of the human c-*sis* gene are indicated by stippled boxes. Exons are indicated by ■. □ indicates the genomic probe that was used for DH site mapping. The horizontal bracket indicates the region of DNase-I-hypersensitivity (see also Fig. 8). H, *Hin*dIII; B, *Bam*HI; E, *Eco*RI; K, *Kpn*I; X, *Xho*I; P, *Pvu*II. (**B**) Inserts from the genomic c-*sis* gene subclones depicted in Figure 5A were examined by Southern blot analysis using the alternative exon 1a as a probe. (**C**) Genomic subclone pAO70 was examined in more detail by Southern blot analysis of restriction enzyme digestions using the alternative exon 1a as a probe.

```
         .         .         .         .        50         .         .         .         .       100         .         .
cagctgcgctggcttccaggagggcgagtcccactgtcacgtgacgcgtctggcc [tcagcacacttcttccgggaaagagtgaagggccccactgccctttgccatccagcttcctctgg

         .         .       150         .         .         .         .       200         .         .         .         .
ctttgctaatggccctaggggcgaggagaccaactgctggaatcccagagccctggaggtgtgcaagggcaggtcaaacagaatttggaggatctggtgcaagagccaggaagagagaga

      *250         .         .         . ↓       .       300         .         .         .         .       350         .
gagagaGTGTGTGTGTGTGTGCGCGCGCATCTGA] GAGAGAGAGAGAGAGAGAGAGACTGACTGAGCAGGAATGGTGAGATGTTTATCATGGGCCTCgtaagtacctctccacgtcttgtc

         .         .         .       400         .         .         .         .       450         .         .         .
ttcccctccccacattgaggagcctcttctgtgacaactcttcctatgttctggtttatttcattgtttattacctgctttctctactggagtgtcaaccccattagagagctttcctcc

         .       500         .         .
tggtccccactttagaacagtgccagagcctgtggacaatcagtaact
```

Figure 6. Nucleotide sequence of the alternative c-*sis* exon 1a and its flanking regions. The sequence starts at the *Pvu*II site immediately upstream of the alternative exon 1a (see also Fig. 5A). Exon 1a sequences are in upper case. The presumed transcription initiation site of the 2.6 kb c-*sis* mRNA is indicated by an asterisk. The 5′ end of cDNA clone pCORD1 is indicated by an arrow. ATG codons in exon 1a are underlined. The consensus translation initiation site according to Kozak (42) is underlined twice. The 0.22 kb *Dde*I fragment that displays promoter activity in a transient reporter gene assay is indicated by brackets (see also Fig. 7).

sponds with the 5′ end of the 2.6 kb mRNA indicated that the transcript contains the regular c-*sis* exons 2–5. By Northern blot analysis with exon-specific probes we could show that it also contains sequences derived from c-*sis* exon 6 and from the 5′ and 3′ parts of exon 7. Primer extension analysis showed that the alternative exon 1a is 90 bp long, which in combination with the regular exons 2–7 fully accounts for the length of the 2.6 kb transcript. Thus, the only difference between the 3.5 and 2.6 kb c-*sis* transcripts is the origin of their first exon sequences.

The alternative c-*sis* expression pathway that is followed in JEG-3 cells may have important qualitative and quantitative consequences for the gene product. The 5′-untranslated sequence of the 2.6 kb c-*sis* mRNA is maximally 81 nt long, has a GC content of only 52% and is not expected to adopt extensive

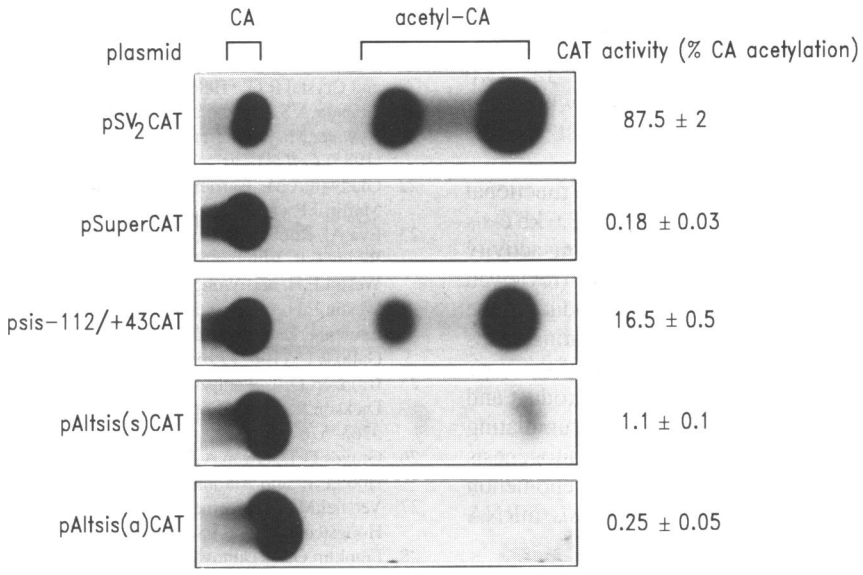


Figure 7. Transient reporter gene analysis of the regular and alternative c-*sis* gene promoter in JEG-3 cells. Cells were electropermeated with a mixture of pCH110 and the indicated CAT constructs as described in Materials and Methods. Aliquots of lysate normalized to mean β-galactosidase activity were tested for CAT activity. CA, chloramphenicol; acetyl-CA, 1- and 3-acetyl-chloramphenicol. All data are means from two to four independent electropermeations.

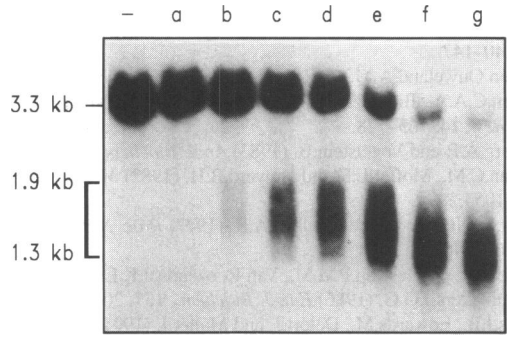


Figure 8. Fine mapping of the DH site located in the center of c-*sis* intron 1. Nuclei from JEG-3 cells were treated with increasing amounts of DNase I. Chromosomal DNA (25 µg) was digested with *Kpn*I and subjected to Southern blot analysis. The DNA was hybridized with the genomic probe depicted in Figure 5. DNase I concentrations: (–) no DNase I added; (**a**) 42 U/ml; (**b**) 84 U/ml; (**c**) 168 U/ml; (**d**) 294 U/ml; (**e**) 462 U/ml; (**f**) 672 U/ml; (**g**) 924 U/ml.

secondary structures. Therefore, as was previously predicted for the 2.8 kb c-*sis* mRNA expressed in vascular endothelial cells (29), the absence of the GC-rich leader sequence may result in the synthesis of significant levels of protein product, even from relatively low levels of mRNA. Much more importantly, translation from the 2.6 kb c-*sis* mRNA may generate a novel PDGF B-like protein product. Exon 1a contains three ATG codons, of which the first one would initiate a peptide of only 20 amino acids. The third ATG codon, which is in frame with the code for the PDGF B chain precursor that continues in exon 2, has the best match with the consensus translation start site according to Kozak (42). Initiation of translation at the third ATG codon is quite possible given the fact that translation of the 3.5 kb c-*sis* mRNA initiates even at the fourth ATG codon, which also meets the criteria of the Kozak consensus sequence (34). The regular exon 1 codes for the signal peptide of the PDGF B chain precursor, which is essential for its translocation to the ER. Translation from the third ATG codon of the 2.6 kb mRNA would generate a polypeptide containing five acidic residues among its first 20 amino acids. Since this acidic N-terminus does not constitute a signal peptide, we predict that the alternative polypeptide is synthesized at 'free' cytoplasmic ribosomes. The regular PDGF B chain precursor dimerizes in the ER and is proteolytically processed in the Golgi system yielding a 24–30 kDa dimeric protein that is secreted or remains membrane associated (13–17). In contrast, the product of the 2.6 kb c-*sis* mRNA species is expected to be a monomeric 25 kDa polypeptide that is initially located in the cytoplasm. The PDGF B chain precursor has been shown to contain a functional nuclear localization signal (NLS) in its exon 6-derived sequence (47). The NLS is usually masked, due to the rapid translocation of the PDGF B chain precursor into the ER, and is subsequently removed by proteolytic processing in the ER. In the absence of a signal peptide, the NLS may target the product of the 2.6 kb c-*sis* mRNA to the cell nucleus.

The alternative exon 1a is preceded by a functional promoter that coincides exactly with a very pronounced DH site in JEG-3 cells. DH sites are thought to represent nucleosome-free regions that often correspond with sites of specific DNA–protein interactions (48). The DH site at the alternative promoter extends over 600 bp, which indicates that three nucleosomes are absent or disrupted. The nuclease S1 and primer extension analyses showed that the 2.6 kb c-*sis* mRNA extends 90 nt upstream of the exon 2-derived sequence, which places the transcription initiation site 90 bp upstream of the 3′ end of exon 1a. The transcription start site coincides with the switch from a $(GA)_7$-repeat to a $(GPy)_{11}$-repeat (Fig. 6) and is not preceded by a consensus TATA box. A $(GA)_{12}$-repeat is located within exon 1a and starts at 64 bp upstream of the 3′ end of the exon. Recently, binding of a GAGA

transcription factor was shown to disrupt nucleosomes at a heat-shock promoter (49). A similar mechanism may be responsible for the generation of the DH site at exon 1a. Two DNase-I-footprints could easily be distinguished within the nucleosome-free region (Fig. 8), which may reflect specific DNA–protein interactions at the alternative promoter.

We are currently focusing on the structural and functional characterization of the putative protein product of the 2.6 kb c-*sis* mRNA. The regular c-*sis* gene product has transforming activity (50–52) and is associated with several tumor species (reviewed in 43). Whether the putative PDGF B-like protein product of the 2.6 kb c-*sis* mRNA has transforming activity is questionable, as it was shown that mutagenesis of the signal peptide of the v-*sis* gene resulted in a nuclear localization of the protein product and a loss of transforming activity (53). There is accumulating evidence that some growth factors exert their role in mitogenesis partly in the nucleus (e.g. 54,55). Whether a similar phenomenon holds true for the putative product of the 2.6 kb c-*sis* mRNA remains to be determined.

ACKNOWLEDGEMENT

We are grateful to Dr H. J. Dodemont for critically reading the manuscript.

REFERENCES

1 Waterfield,M.D., Scrace,G.T., Whittle,N., Stroobant,P., Johnsson,A., Wasteson,Å., Westermark,B., Heldin,C.-H., Huang,J.S. and Deuel,T.F. (1983) *Nature*, **304**, 35–39.
2 Doolittle,R.F., Hunkapiller,M.W., Hood,L.E., Devare,S.G., Robbins,K.C., Aaronson,S.A. and Antoniades, H.N. (1983) *Science*, **221**, 275–277.
3 Josephs,S.F., Guo,C., Ratner,L. and Wong Staal,F. (1984) *Science*, **223**, 487–491.
4 Johnsson,A., Heldin,C.H., Wasteson,A., Westermark,B., Deuel,T.F., Huang,J.S., Seeburg,P.H., Gray,A., Ullrich,A., Scrace,G., Stroobant,P. and Waterfield,M.D. (1984) *EMBO J.*, **3**, 921–928.
5 Chiu,I.-M., Reddy,E.P., Givol,D., Robbins,K.C., Tronick,S.R. and Aaronson,S.A. (1984) *Cell*, **37**, 123–129.
6 Dalla Favera,R., Gallo,R.C., Giallongo,A. and Croce,C.M. (1982) *Science*, **218**, 686–688.
7 Swan,D.C., McBride,O.W., Robbins,K.C., Keithly,D.A., Reddy,E.P. and Aaronson,S.A. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4691–4695.
8 Rao,C.D., Igarashi,H., Chiu,I.M., Robbins,K.C. and Aaronson,S.A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 2392–2396.
9 Van den Ouweland,A.M., van Groningen,J.J., Schalken,J.A., van Neck,H.W., Bloemers,H.P.J. and Van de Ven,W.J. (1987) *Nucleic Acids Res.*, **15**, 959–970.
10 Ratner,L., Thielan,B. and Collins,T. (1987) *Nucleic Acids Res.* **15**, 6017–6036.
11 Rao,C.D., Pech,M., Robbins,K.C. and Aaronson,S.A. (1988) *Mol. Cell. Biol.*, **8**, 284–292.
12 Ratner,L. (1989) *Nucleic Acids Res.*, **17**, 4101–4115.
13 Graves,D.T., Owen,A.J., Williams,S.R. and Antoniades,H.N. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4636–4640.
14 Igarashi,H., Rao,C.D., Siroff,M., Leal,F., Robbins,K.C. and Aaronson,S.A. (1987) *Oncogene*, **1**, 79–85.
15 Bywater,M., Rorsman,F., Bongcam Rudloff,E., Mark,G., Hammacher,A., Heldin,C.H., Westermark,B. and Betsholtz,C. (1988) *Mol. Cell. Biol.*, **8**, 2753–2762.
16 Östman,A., Rall,L., Hammacher,A., Wormstead,M.A., Coit,D., Valenzuela,P., Betsholtz,C., Westermark,B. and Heldin,C.H. (1988) *J. Biol. Chem.*, **263**, 16 202–16 208.
17 Östman,A., Thyberg,J., Westermark,B. and Heldin,C.H. (1992) *J. Cell Biol.*, **118**, 509–519.
18 Barrett,T.B., Gajdusek,C.M., Schwartz,S.M., McDougall,J.K. and Benditt,E.P. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 6772–6774.
19 Shimokado,K., Raines,E.W., Madtes,D.K., Barrett,T.B., Benditt,E.P. and Ross,R. (1985) *Cell*, **43**, 277–286.
20 Martinet,Y., Bitterman,P.B., Mornex,J.F., Grotendorst,G.R., Martin,G.R. and Crystal,R.G. (1986) *Nature*, **319**, 158–160.
21 Goustin,A.S., Betsholtz,C., Pfeifer-Ohlsson,S., Persson,H., Rydnert,J., Bywater,M., Holmgren,G., Heldin,C.H., Westermark,B. and Ohlsson,R. (1985) *Cell*, **41**, 301–312.
22 Gladwin,A.M., Carrier,M.J., Beesley,J.E., Lelchuk,R., Hancock,V. and Martin,J.F. (1990) *Br. J. Haemat.*, **76**, 333–339.
23 Eva,A., Robbins,K.C., Andersen,P.R., Srinivasan,A., Tronick,S.R., Reddy,E.P., Ellmore,N.W., Galen,A.T., Lautenberger,J.A., Papas,T.S., Westin,E.H. and Wong-Staal,F. (1982) *Nature*, **295**, 116–119.
24 Westin,E.H., Wong-Staal,F., Gelmann,E.P., Dalla Favera,R., Papas,T.S., Lautenberger,J.A., Eva,A., Reddy,E.P., Tronick,S.R., Aaronson,S.A. and Gallo,R.C. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 2490–2494.
25 Bronzert,D.A., Pantazis,P., Antoniades,H.N., Kasid,A., Davidson,N., Dickson,R.B. and Lippman, M.E. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5763–5767.
26 Graves,D.T., Owen,A.J., Barth,R.K., Tempst,P., Winoto,A., Fors,L., Hood,L.E. and Antoniades,H.N. (1984) *Science*, **226**, 972–974.
27 Versnel,M.A., Hagemeijer,A., Bouts,M.J., van der Kwast,T.H. and Hoogsteden,H.C. (1988) *Oncogene*, **2**, 601–605.
28 Franklin,G.C., Donovan,M., Adam,G.I., Holmgren,L., Pfeifer-Ohlsson,S. and Ohlsson,R. (1991) *EMBO J.*, **10**, 1365–1373.
29 Fen,Z. and Daniel,T.O. (1991) *Oncogene*, **6**, 953–959.
30 Dirks,R.P.H., Jansen,H.J., Gerritsma,J., Onnekink,C. and Bloemers,H.P.J. (1993) *Eur. J. Biochem.*, **211**, 509–519.
31 Auffray,C. and Rougeon,F. (1980) *Eur. J. Biochem.*, **107**, 303–314.
32 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning. A Laboratory Manual* (2nd edn). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
33 Van den Ouweland,A.M.W., Breuer,M.L., Steenbergh,P.H., Schalken,J.A., Bloemers,H.P.J. and Van de Ven,W.J.M. (1985) *Biochim. Biophys. Acta*, **825**, 140–147.
34 Van den Ouweland,A.M.W., Roebroek,A.J.M., Schalken,J.A., Claesen,C.A.A., Bloemers,H.P.J. and Van de Ven,W.J.M. (1986) *Nucleic Acids Res.*, **14**, 765–778.
35 Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13.
36 Gorman,C.M., Moffat,L.F. and Howard,B.H. (1982) *Mol. Cell. Biol.*, **2**, 1044–1054.
37 Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
38 Van Leen,R.W., Kastrop,P.M.M., Van Roozendaal,K.E.P. and Schoenmakers,J.G.G. (1986) *Eur. J. Biochem.*, **157**, 203–208.
39 Dumas,J.B., Edwards,M., Delort,J. and Mallet,J. (1991) *Nucleic Acids Res.*, **19**, 5227–5232.
40 Edlund,T., Walker,M.D., Barr,P.J. and Rutter,W.J. (1985) *Science*, **230**, 912–916.
41 Dirks,R.P.H., Jansen,H.J., Onnekink,C., De Jonge,R.J. and Bloemers,H.P.J. (1993) *Eur. J. Biochem.*, **216**, 487–495.
42 Kozak,M. (1991) *J. Cell Biol.*, **115**, 887–903.
43 Ross,R., Raines,E.W. and Bowen-Pope,D.F. (1986) *Cell*, **46**, 155–169.
44 Kaetzel,D.M., Coyne,D.W. and Fenstermaker,R.A. (1993) *BioFactors*, **4**, 71–81.
45 Press,R.D., Samols,D. and Goldthwait,D.A. (1988) *Biochemistry*, **27**, 5736–5741.
46 Sariban,E. and Kufe,D. (1988) *Cancer Res.*, **48**, 4498–4502.
47 Lee,B.A., Maher,D.W., Hannink,M. and Donoghue,D.J. (1987) *Mol. Cell. Biol.*, **7**, 3527–3537.
48 Gross,D.S. and Garrard,W.T. (1988) *Annu. Rev. Biochem.*, **57**, 159–197.
49 Tsukiyama,T., Becker,P.B. and Wu,C. (1994) *Nature*, **367**, 525–532.
50 Clarke,M.F., Westin,E., Schmidt,D., Josephs,S.F., Ratner,L., Wong Staal,F., Gallo,R.C. and Reitz,M.S. Jr (1984) *Nature*, **308**, 464–467.
51 Gazit,A., Igarashi,H., Chiu,I.M., Srinivasan,A., Yaniv,A., Tronick,S.R., Robbins,K.C. and Aaronson,S.A. (1984) *Cell*, **39**, 89–97.
52 Josephs,S.F., Ratner,L., Clarke,M.F., Westin,E.H., Reitz,M.S. and Wong Staal,F. (1984) *Science*, **225**, 636–639.
53 Hannink,M. and Donoghue,D.J., (1986) *Mol. Cell Biol.*, **6**, 1343–1348.
54 Kimura,H. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 2165–2169.
55 Wiedlocha,A., Falnes,P.Ø., Madshus,I.H., Sandvig,K. and Olsnes,S. (1994) *Cell*, **76**, 1039–1051.